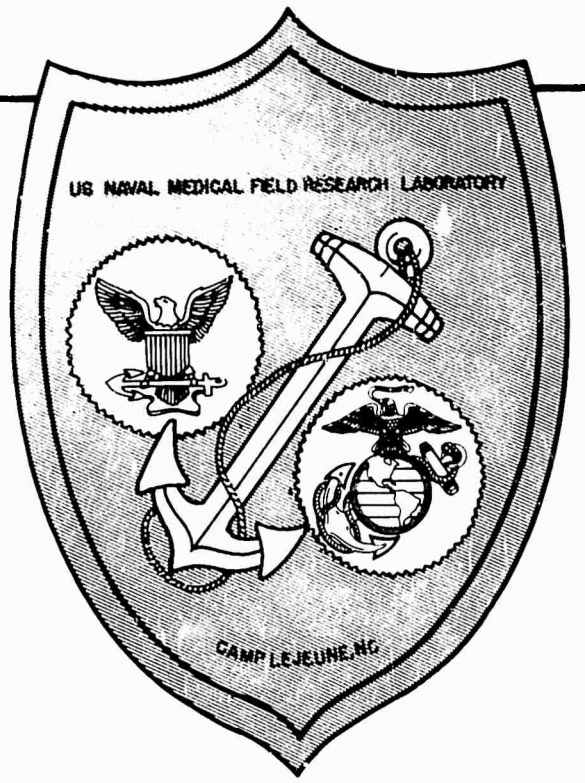


U.S. NAVAL MEDICAL FIELD
RESEARCH LABORATORY
CAMP LEJEUNE, NORTH CAROLINA



Vol. XXI, No. 3

January 1971

# A STUDY OF THE RELIABILITY OF THE USMC PHYSICAL FITNESS TEST

by

Philip J. Rasch, Ph.D.



Bureau of Medicine and Surgery, Navy Department
Work Unit MF12.524.007-8013BA8X 18



15

MCBCL 1375A

Vol. XXI, No. 3 January 1971

# A STUDY OF THE RELIABILITY OF THE USMC PHYSICAL FITNESS TEST

by

Philip J. Rasch, Ph.D.

Physiology Division
NAVAL MEDICAL FIELD RESEARCH LABORATORY
CAMP LEJEUNE, NORTH CAROLINA 28542

Bureau of Medicine and Surgery, Navy Department
Work Unit MF12.524.007-8013BA8X.18



Submitted by:

PHILIP J. RASCH
Chief
Physiology Division

Approved by:

JESSE F. ADAMS
CAPT MC USN
Commanding Officer

## SUMMARY PAGE

### THE PROBLEM

To determine the test-retest reliability of the events in the Marine Corps Physical Fitness Test.

### FINDINGS

The following events had test-retest correlations of $r = 0.80$ or higher: pull-ups, rope climb, bent knee sit-ups, standing broad jump, jump and reach, three-mile run, and 300-yard shuttle run. The following had test-retest correlations of less than $r=0.80$: push-ups, leg lifts in two minutes, bend and thrust in one minute.

### RECOMMENDATIONS

1. Assuming two events are of equal validity, it is usually preferable to use the one with the higher reliability.

2. Troops should be thoroughly trained in a test before it is administered for score.

3. The instructions for the 300-yard run are not standardized and it does not correlate well with the three-mile run.

## ADMINISTRATIVE INFORMATION

Bureau of Medicine and Surgery, Department of the Navy, Work Unit MF12.524.007-8013BA8X, report 18. Interim report. Approved for publication 16 December 1970.

Published by the Naval Medical Field Research Laboratory, Camp Lejeune, North Carolina 28542.

This restriction will be removed and the report may be released on 31 March 1971.

## ABSTRACT

The Marine Corps Physical Fitness Test was administered on a test-retest basis. The following events had test-retest correlations of $r = 0.80$ or higher: pull-ups, rope climb, bent knee sit-ups in two minutes, standing broad jump, jump and reach, three-mile run, and 300-yard shuttle run. The following events had test-retest correlations of less than $r = 0.80$: push-ups, leg lifts in two minutes, bend and thrust in one minute.

## BACKGROUND

If a test battery is to meet minimum acceptable standards, the instructions must include certain technical information, such as test-retest reliability coefficients, validity coefficients or assumptions, intercorrelations of the individual test items, and norms. When the present Marine Corps Physical Fitness Test was devised, no test-retest correlations were computed. It was the purpose of this study to determine these relationships.

## PROCEDURE AND RESULTS

The Marine Corps Physical Fitness Test was administered to two groups of troops on two separate occasions. Seven to 10 days were allowed between tests. During this period, four subjects were lost for one reason or another. Since it was desired to administer every item in the test, on each occasion the actual data collection was spread over a two-day period. The order of administration was kept the same and was as follows:

| Test Days 1 and 3 | Test Days 2 and 4 |
|---|---|
| Push-ups | Pull-ups |
| Rope Climb | Bent Knee Sit-ups in two minutes |
| Leg Lifts in two minutes | Bend and Thrust in one minute |
| Standing Broad Jump | Jump and Reach |
| 300-yard Shuttle Run | Three-mile Run |

The testing program was preceded each day by a light warm-up. A five-minute rest period was observed between each two tests. The uniform was utilities and boots. The subjects were from several different commands and represented a good cross-section of Marine Corps personnel.* Unfortunately, the weather was unsettled, with occasional showers. This made the ground slippery. Undoubtedly, the times for the runs were longer than would have been the case otherwise and possibly the correlations may have been somewhat attenuated.

The usual statistic expressing the reliability of a measurement is the correlation ($r$) between repeated tests. In evaluating the correlation, the interpretations attributed to Clarke[1] were accepted.

| | |
|---|---|
| .95 to .99 | Very high: rarely found among present tests. |
| .90 to .94 | High: equalled by a few of the best tests. |
| .80 to .89 | Fairly high: fairly adequate for individual measurement. |
| .70 to .79 | Rather low: adequate for group measurement but not very satisfactory for individual measurements. |
| Below .70 | Low: entirely inadequate for individual measurements, although useful for group averages and school surveys. |

*The writer is indebted to 1st Lieutenant Thomas Blair Trammell, USMC, Marine Corps Liaison Officer, Naval Medical Field Research Laboratory, for his assistance in securing troops for this study.

For the purposes of this study, a test-retest correlation of $r = 0.80$ was accepted as satisfactory, although this is probably lower than many physical educators would be prepared to accept. For convenience sake, each test item will be considered separately in what follows.

1. Pull-ups

The means of the data for the test-retest scores are shown in Table 1

TABLE 1
*Pull-ups – Test-Retest*
*N = 27*

| First Trial | | Second Trial | | |
|---|---|---|---|---|
| Mean | S.D.† | Mean | S.D. | r |
| 7.6 | 4.39 | 7.7 | 4.78 | 0.92 |

† S.D. = Standard deviation

This reliability correlation is excellent; it is not likely that there are many tests of physical performance which will score as high or higher. The $r = 0.92$ is for all practical purposes identical with the $r = 0.93$ found by Fleishman in a study of 201 recruits at Great Lakes Naval Training Center.[2]

2. Push-ups

The means of the data collected for the push-ups are shown in Table 2.

TABLE 2
*Push-ups – Test-Retest*
*N = 24*

| First Trial | | Second Trial | | |
|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | r |
| 39.0 | 12.3 | 42.5 | 15.5 | 0.77 |

While adequately reliable for group measurements, this correlation is slightly below the acceptable figure. With more experienced troops or a greater number of subjects, it might have reached an acceptable figure. According to Fleishman,[3] pull-ups have a loading of $r = 0.81$ on Dynamic Strength and push-ups have a loading of $r = 0.74$. The coefficient of determination* for the two events is thus 0.65 for pull-ups versus 0 55 for push-ups. This suggests that use of pull-ups as a standard is preferable to the use of push-ups

3. Rope Climb

The means of the data collected in this event are presented in Table 3.

TABLE 3
*Rope Climb (in seconds)*
*N = 29*

| First Trial | | Second Trial | | |
|---|---|---|---|---|
| Mean | S.D | Mean | S.D. | r |
| 12.5 | 4.02 | 12.4 | 4.32 | 0.81 |

This correlation is fairly high and is adequate for individual measurements, although barely so.

*The coefficient of determination ($d$) is computed as $d = r^2$. When multiplied by 100, it gives the percentage of the variance in one factor that is associated with the variance in the other.

4. Bent Knee Sit-ups in Two Minutes

The means of the data collected in this test item are shown in Table 4.

TABLE 4
*Bent Knee Sit-ups in Two Minutes*
*N = 27*

| First Trial | | Second Trial | | |
|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | *r* |
| 45.9 | 12.28 | 47.0 | 10.04 | 0.83 |

The reliability of this test is thus fairly adequate for individual measurements. However, it should be noted that a recent publication from this laboratory[4] raises a question as to its validity.

Observation revealed that very few of the subjects knew how to pace themselves in this event. Almost invariably they started out too fast and then had to rest at some period during the two minutes. This suggests that a considerably higher test-retest correlation would be found in the case of subjects who are more experienced in taking this test.

These subjects, and a number of others, had also participated in a study of the test recommended by the International Committee on the Standardization of Physical Fitness Tests. This battery includes a 30-second bent knee sit-up test. Scores in the two events were correlated (Table 5), using the first two-minute test as the sources of the scores for that item, to determine whether the shorter test would be satisfactory for Marine Corps use. This is significant at the .01 level and indicates that there is a strong relationship between the two tests. However, the correlation is too low to justify substituting the 30-second test for the two-minute event.

TABLE 5
*Correlation of 30-second Bent Knee Sit-ups with 2-minute Bent Knee Sit-ups*
*N = 50*

| 30-sec Sit-ups | | 2-min Sit-ups | | |
|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | *r* |
| 24 | 4.54 | 47 | 11.76 | 0.40 |

5. Leg Lifts in Two Minutes

The means of the data collected during this test are shown in Table 6.

This correlation is too low to be satisfactory for individual measurement, although of some value for group surveys. There is also some question as to the validity of this event when used to test the combat fitness of Marine troops.[4] Further study of this test item is indicated.

TABLE 6
*Leg Lifts in Two Minutes*
*N = 24*

| First Trial | | Second Trial | | |
|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | *r* |
| 41.1 | 11.8 | 46.6 | 12.6 | 0.62 |

6. Bend and Thrust in One Minute

The means of the data collected in this event are shown in Table 7.

The reliability of this event is definitely unsatisfactory. In addition, there is a serious question

TABLE 7
*Bend and Thrust in One Minute*
*N = 26*

| First Trial | | Second Trial | | |
|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | *r* |
| 29.5 | 4.65 | 33.7 | 6.80 | 0.56 |

regarding its validity. The writer is unable to find any investigations of this in the literature. Observation indicates that it contains factors of agility, speed, coordination, and cardiorespiratory endurance, although the relative proportions of each have not been determined. It is recommended that further studies of this test be undertaken to determine whether it has any place in the Marine Corps Physical Fitness Test.

7. Standing Broad Jump

In the standing broad jump* the trooper is given three attempts. The mean scores and intercorrelations of the first test on each occasion are displayed in Table 8.

TABLE 8

*First Test – Standing Broad Jump (in inches)*
*N = 33*

| First Trial | | Second Trial | | Third Trial | | |
|---|---|---|---|---|---|---|
| Mean | S.D. | Mean | S.D | Mean | S.D. | r |
| 81 2 | 9.03 | 82.9 | 8.97 | | | 0.87 |
| 81.2 | 9.03 | | | 82.8 | 10.00 | 0.86 |
| | | 82.9 | 8.97 | 82.8 | 10.00 | 0.85 |

The same data for the second test are set forth in Table 9.

TABLE 9

*Second Test – Standing Broad Jump (in inches)*
*N = 28*

| Fourth Trial | | Fifth Trial | | Sixth Trial | | |
|---|---|---|---|---|---|---|
| Mean | S D. | Mean | S.D. | Mean | S.D | r |
| 80.0 | 8.75 | 82.5 | 9.48 | | | 0.84 |
| 80.0 | 8.75 | | | 82.3 | 10.68 | 0.84 |
| | | 82.5 | 9.48 | 82.3 | 10.68 | 0.88 |

These figures compare quite satisfactorily with those of Fleishman. He found a mean of 82.94 inches, and S.D. of 8.73 inches, and a test-retest coefficient of $r = 0.90$ in his naval subjects.[2]

A comparison between the distances achieved on the first and second occasions of testing is given in Table 10.

TABLE 10

*Standing Broad Jump – First Test vs Second Test (in inches)*
*N = 28*

| First Trial | | Second Trial | | Third Trial | | Fourth Trial | | Fifth Trial | | Sixth Trial | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | S D | Mean | S D | Mean | S.D | Mean | S.D. | Mean | S D | Mean | S.D. | r |
| 81 2 | 9 03 | | | | | 80 0 | 8.75 | | | | | 0.80 |
| | | 82 9 | 8 97 | | | | | 82.5 | 9.48 | | | 0.85 |
| | | | | 82.8 | 10.0 | | | | | 82.3 | 10.68 | 0.88 |

These data are consistent in displaying a satisfactorily high test-retest correlation.

*Physical educators now term this event the standing long jump It would be desirable that Marine Corps usage be brought into agreement.

## 8. Jump and Reach

In the jump and reach,* the trooper is given three attempts. The mean scores and intercorrelations on the first test were as shown in Table 11.

TABLE 11

*First Test – Jump and Reach (in inches)*

*N = 31*

| First Trial | | Second Trial | | Third Trial | | |
|---|---|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | Mean | S.D. | *r* |
| 17.3 | 3.05 | 17.6 | 3.21 | | | 0.91 |
| 17.3 | 3.05 | | | 18.0 | 3.30 | 0.91 |
| | | 17.6 | 3.21 | 18.0 | 3.30 | 0.94 |

The mean scores and intercorrelations on the second set of jumps are shown in Table 12.

TABLE 12

*Second Test – Jump and Reach (in inches)*

*N = 27*

| Fourth Trial | | Fifth Trial | | Sixth Trial | | |
|---|---|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | Mean | S.D. | *r* |
| 17.7 | 3.10 | 18.0 | 2.96 | 17.9 | 2.81 | 0.92 |
| 17.7 | 3.10 | | | 17.9 | 2.81 | 0.86 |
| | | 18.0 | 2.96 | 17.9 | 2.81 | 0.86 |

We can conclude from this that when three trials are given, the test-retest correlations range from fairly adequate to high. The reliability figures are so close to those of the standing broad jump that there would seem to be little to choose between them in this respect, although the latter appears to be more stable over a period of time. The data compare well with those of Fleishman, who found a mean of 18.43 inches, and S.D. of 2.66 inches, and a test-retest correlation of $r = 0.90$ in his study of naval recruits.[2]

A comparison between the heights achieved on the first and second occasions of testing is given in Table 13.

TABLE 13

*Jump and Reach – First Test vs Second Test (in inches)*

*N = 27*

| First Trial | | Second Trial | | Third Trial | | Fourth Trial | | Fifth Trial | | Sixth Trial | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. | *r* |
| 17.1 | 3.1 | | | | | 17.7 | 3.1 | | | | | 0.69 |
| | | 17.3 | 3.11 | | | | | 18.0 | 2.96 | | | 0.71 |
| | | | | 17.7 | 3.03 | | | | | 17.9 | 2.81 | 0.73 |

*The jump and reach test is generally known to physical educators as the vertical jump to make the nomenclature consistent with the long jump (broad jump). It would be desirable that the Marine Corps terminology be brought into agreement.

When the mean scores made on two separate occasions are compared, the correlation becomes unsatisfactory for individual measurements. In all probability, this is due to small changes in stance and similar factors. Martin and Stull[5] have shown that knee angle, lateral foot spacing, and anterior-posterior foot space exert independent effects on vertical jumping performance. Under a laboratory situation, it would be possible to have pre-jump positions indicated, but this appears impractical for use in the field.

9. Three-mile Run

The three-mile run is apparently peculiar to the Marine Corps Physical Fitness Test. At least the writer has not been able to find any reference to it in the literature. The means of the data are shown in Table 14.

TABLE 14
*Three-mile Run (in minutes)*
*N = 24*

| First Trial | | Second Trial | | |
|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | *r* |
| 26.9 | 2.90 | 26.7 | 3.18 | 0.84 |

This event thus has a satisfactorily high reliability coefficient.

10. 300-yard Shuttle Run

Change 2 to Marine Corps Order 6100.3E provides that under certain conditions the 300-yard shuttle run may be substituted for the three-mile run. Rather curiously, the change does not give any information on how the shuttle run shall be conducted. In the absence of any specific instructions, we laid off two lime lines 50 yards apart. The subjects started with their forward foot on one line and were required to get both feet across the second line before returning to the starting line. The men trotted through the course to familiarize themselves with it and to warm up before running the event for a score. This same procedure was followed during the retest. The means of the data are shown in Table 15.

This test item has a satisfactory reliability, but the fact that it was necessary to lay out the course on a grassy surface which was constantly kept wet by the intermittent showers experienced during the testing almost surely attenuated the correlations.

TABLE 15
*300-yard Shuttle Run (in seconds)*
*N = 24*

| First Trial | | Second Trial | | |
|---|---|---|---|---|
| Mean | S.D. | Mean | S.D. | *r* |
| 58.6 | 4.56 | 58.0 | 3.79 | 0.80 |

Since this is a new item, intercorrelations were computed with the other items in the test battery to determine whether or not it is orthogonal * to the other items in the battery. Also, it was desired to determine how it correlates with the three-mile run. It is obvious that an assumption has been made that the 300-yard shuttle run has a high correlation with the three-mile run and a low correlation with every other item in the Physical Fitness Test battery. The writer knows of no evidence to support such a hypothesis. It was assumed that the second run might be the more stable, the first serving to practice the requisite skills. The

*Comparisons are said to be orthogonal when they are statistically independent; i.e., the information provided by one is unrelated to the information provided by the other. This is shown by a low correlation between the two test items.

second trial was also used in the case of the other events. The intercorrelations thus obtained are shown in Table 16.

It will be noted that the shuttle run is orthogonal to all other events in the Marine Corps Physical Fitness Test and has a very low correlation with the three-mile run. Accordingly, the assumption that it can be substituted for that event is incorrect.

TABLE 16

*Intercorrelations of 300-yard Shuttle Run with Other Items in the Physical Fitness Test*

| Event | *r* | N |
|---|---|---|
| Pull-ups | 0.453 | 27 |
| Push-ups | 0.446 | 29 |
| Rope Climb | 0.232 | 29 |
| Bent Knee Sit-ups in 2 minutes | 0.071 | 27 |
| Leg Lifts in 2 minutes | 0.464 | 29 |
| Bend and Thrust in 1 minute | 0.075 | 27 |
| Standing Broad Jump | 0.508 | 29 |
| Jump and Reach | 0.433 | 28 |
| 3-mile Run | 0.259 | 24 |

## DISCUSSION

One of the assumptions made in undertaking this investigation was that all subjects would be thoroughly familiar with the test events as a result of having repeatedly taken the Marine Corps Physical Fitness Test. In retrospect, this assumption does not appear to have been justifiable. Several of our subjects reported that they had very little experience with this test and it was evident from watching them that they were unaware of how to pace themselves in the timed events. McGraw and McClenney[6] have reported that boys administered push-ups, sit-ups, and pull-ups improved significantly in performance during four trials on each of the tests. In general, their findings were confirmed by Baumgartner,[7] who observed that in some events the individual learned to perform the test better on each attempt over a period of several trials.

It was evident that most of our subjects had not had sufficient experience with these tests to permit the requisite stabilization of test scores to occur. If the Corps wishes to get a true picture of the fitness of its troops, it must administer its Physical Fitness Test for score only after the troops have been administered the test for practice on several occasions and have become familiar with their own capabilities and with the technique required by each event.

It also seems clear that a similar procedure with a larger number of subjects than were available for this study will have to be followed before an accurate determination test-retest correlation could be made. Due to the weather routinely experienced at Camp Lejeune, it would be desirable that a test program of this sort be conducted at Camp Pendleton.

In the case of the 300-yard shuttle run, it seems peculiar that the instructions provide "This event will not be scored by age groups."[8] It is a matter of common experience that as men become older they also become less agile. Considering this and the fact that all other events in the Marine Corps Physical Fitness Test are scored by age groups, this restriction simply does not appear reasonable. It is recommended that a study be made of the validity of this restriction.

## CONCLUSIONS

The results of this limited study were as follows:

1. Assuming a test-retest correlation of $r = 0.80$ as acceptable,

   a. The following test items were found satisfactory in this respect:

      (1) Pull-ups
      (2) Rope climb
      (3) Bent knee sit-ups in two minutes
      (4) Standing broad jump
      (5) Three-mile run
      (6) 300-yard shuttle run
      (7) Jump and reach

   b. The following test items were found unsatisfactory in this respect:

      (1) Push-ups
      (2) Leg lifts in two minutes
      (3) Bend and thrust in one minute

2. The 300-yard shuttle run as administered in this test has a negligible correlation with the three-mile run and cannot be used as a substitute for it.

3. The 300-yard shuttle run is satisfactorily orthogonal to the other test events.

4. The apparent assumption that scores in the 300-yard shuttle run are not affected by age is highly questionable. The validity of this assumption should be investigated.

5. It is recommended that this investigation be repeated, using larger numbers, better trained subjects, and under more favorable weather conditions, to obtain more accurate data.

## REFERENCES

1. Weber, Jerome C. and Lamb, David R. *Statistics and Research in Physical Education.* Saint Louis: The C.V. Mosby Company, 1970, pp. 182-183.

2. Fleishman, Edwin A. *The Structure and Measurement of Physical Fitness.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964, p. 59.

3. *Ibid.*, p. 64.

4. Rasch, Philip J. *Evaluation of the ICSPFT Test for Possible Marine Corps Use.* NMFRL report, Vol. 21, No. 2, January 1971.

5. Martin, Thomas P. and Scull, G. Alan. Effects of Various Knee Angle and Foot Spacing Combinations on Performance in the Vertical Jump. *Res. Quart.* 40:324-331, May 1969.

6. McGraw, Lynn W. and McClenney, Byron N. Reliability of Fitness Strength Tests. *Res. Quart.* 36:289-295, October 1965.

7. Baumgartner, Ted A. Stability of Physical Performance Test Scores. *Res. Quart.* 40:257-261, May 1969.

8. Marine Corps Order 6100.3E, Change 2, 17 February 1970, p. 12.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Naval Medical Field Research Laboratory
Camp Lejeune, North Carolina 28542

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

A STUDY OF THE RELIABILITY OF THE USMC PHYSICAL FITNESS TEST

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Interim report

5. AUTHOR(S) (First name, middle initial, last name)

Philip J. Rasch, PhD

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| January 1971 | 15 | 8 |

| 8a. CONTRACT OR GRANT NO | 9a. ORIGINATOR'S REPORT NUMBER(S) |
|---|---|
| b. PROJECT NO MF12.524 | MF12.524.007-8013BA8X.18 |
| c. Task No. MF12.524.007 | 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) |
| d. Work Unit No. MF12.524.007-8013BA8X | Vol. 21, No. 3, Jan 71 |

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY |
|---|---|
| | Bureau of Medicine and Surgery<br>Navy Department<br>Washington, D. C. 20390 |

13. ABSTRACT

The Marine Corps Physical Fitness Test was administered on a test-retest basis. The following events had test-retest correlations of $r = 0.80$ or higher: pull-ups, rope climb, bent knee sit-ups in two minutes, standing broad jump, jump and reach, three-mile run, and 300-yard shuttle run. The following events had test-retest correlations of less than $r = 0.80$: push-ups, leg lifts in two minutes, bend and thrust in one minute. (U)

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0102-014-6700

UNCLASSIFIED

Security Classification